**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/00 — A1

(11) International Publication Number: **WO 97/06245**

(43) International Publication Date: 20 February 1997 (20.02.97)

(21) International Application Number: PCT/GB96/01906

(22) International Filing Date: 5 August 1996 (05.08.96)

(30) Priority Data:
9516059.4 4 August 1995 (04.08.95) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: DAVIES, Kay, Elizabeth [GB/GB]; University of Oxford, Genetics Laboratory, Dept. of Biochemical, South Parks Road, Oxford OX1 3QU (GB). THEODOSIOU, Aspasia [GR/GB]; University of Oxford, Genetics Laboratory, Dept. of Biochemistry, South Parks Road, Oxford OX1 3QU (GB).

(74) Agent: BRASNETT, Adrian, Hugh; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5LX (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: MAP KINASE PHOSPHATASE GENE AND USES THEREOF

(57) Abstract

A novel gene is disclosed that encodes a novel dual specificity threonine-tyrosine phosphatase. Is is expressed highly in the brain and may have utility in investigating signal transduction mechanisms in brain and muscle. The isolation of the gene allows one to probe for any mutations that may occur, for example extensions of the gene.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| **AM** | Armenia | **GB** | United Kingdom | **MW** | Malawi |
| **AT** | Austria | **GE** | Georgia | **MX** | Mexico |
| **AU** | Australia | **GN** | Guinea | **NE** | Niger |
| **BB** | Barbados | **GR** | Greece | **NL** | Netherlands |
| **BE** | Belgium | **HU** | Hungary | **NO** | Norway |
| **BF** | Burkina Faso | **IE** | Ireland | **NZ** | New Zealand |
| **BG** | Bulgaria | **IT** | Italy | **PL** | Poland |
| **BJ** | Benin | **JP** | Japan | **PT** | Portugal |
| **BR** | Brazil | **KE** | Kenya | **RO** | Romania |
| **BY** | Belarus | **KG** | Kyrgystan | **RU** | Russian Federation |
| **CA** | Canada | **KP** | Democratic People's Republic of Korea | **SD** | Sudan |
| **CF** | Central African Republic | | | **SE** | Sweden |
| **CG** | Congo | **KR** | Republic of Korea | **SG** | Singapore |
| **CH** | Switzerland | **KZ** | Kazakhstan | **SI** | Slovenia |
| **CI** | Côte d'Ivoire | **LI** | Liechtenstein | **SK** | Slovakia |
| **CM** | Cameroon | **LK** | Sri Lanka | **SN** | Senegal |
| **CN** | China | **LR** | Liberia | **SZ** | Swaziland |
| **CS** | Czechoslovakia | **LT** | Lithuania | **TD** | Chad |
| **CZ** | Czech Republic | **LU** | Luxembourg | **TG** | Togo |
| **DE** | Germany | **LV** | Latvia | **TJ** | Tajikistan |
| **DK** | Denmark | **MC** | Monaco | **TT** | Trinidad and Tobago |
| **EE** | Estonia | **MD** | Republic of Moldova | **UA** | Ukraine |
| **ES** | Spain | **MG** | Madagascar | **UG** | Uganda |
| **FI** | Finland | **ML** | Mali | **US** | United States of America |
| **FR** | France | **MN** | Mongolia | **UZ** | Uzbekistan |
| **GA** | Gabon | **MR** | Mauritania | **VN** | Viet Nam |

# MAP KINASE PHOSPHATASE GENE AND USES THEREOF

This invention relates to two genes encoding novel proteins that possesses threonine-tyrosine phosphatase characteristics, to the proteins themselves and methods for their recombinant production. These genes are located in the cytoplasm which is a novel feature of this class of phosphatase genes.

Protein tyrosine phosphatases (PTPs) are a growing family of enzymes which play an important role, together with protein tyrosine kinases, in many cellular processes such as cell division, proliferation and differentiation[1-3]. The PTP family can be sub-divided according to structural features which determine whether they are transmembrane or cytoplasmically located. All PTPs contain a catalytic domain consisting of a highly conserved active site with the consensus sequence [I/V]HCXAGXXR[S/T]G (where X represents any amino acid). The regions flanking the catalytic domain of the PTPs are diverse and consist of sequences which appear to target the PTPs to specific cellular locations.

Amongst the superfamily of tyrosine phosphatases are a sub-family of dual specificity phosphatases, so-called because they can dephosphorylate substrates which are phosphorylated on both serine and threonine as well as tyrosine residues. Several of these enzymes can dephosphorylate and deactivate MAP kinase (Mitogen activated protein kinases). Genes for some of these MAP kinase phosphatases are known.

The mechanism by which extracellular signals for growth and differentiation are transmitted to the nucleus to alter gene expression is the subject of much current investigation. In many cases, the transduction of these signals requires the activities of key enzymes known generally as MAP kinases. MAP kinase pathways have been implicated in a large number of signal transduction pathways. For instance, the activation of MAP kinases has been observed during growth factor stimulation of DNA synthesis and during differentiation, secretion and stimulation of glycogen synthesis[4]. MAP kinase has been shown to

phosphorylate and activate effector substrates such as the transcription factors c-jun and elk-1. Known MAP kinases, and the pathways in which they are involved, have been reviewed[5].

Map kinase is activated by phosphorylation of threonine and tyrosine by a dual specificity kinase, "MAP kinase kinase". This kinase is in turn activated by phosphorylation by "MAP kinase kinase kinase", one form of which is the proto-oncogene c-raf. The activation of c-raf is not fully understood at present but apparently there is a requirement for an interaction with GTP-bound p21 ras protein[6].

The full picture of how MAP kinase pathways are switched off is not yet clear. Down-regulation of MAP kinase activity by dephosphorylation is likely to be of key importance. The human gene CL100[7] and its murine homologue 3CH134[8] were originally discovered as genes whose transcription was stimulated by growth factors, oxidative stress and heat shock. Subsequently, they were shown to encode polypeptides that have both serine/threonine and tyrosine phosphatase activity[9,10] on MAP kinase. This removal of phosphate from both threonine and tyrosine on MAP kinase is unusual. When expressed *in vitro*[6] the CL100[10] gene product has been shown to be very specific for MAP kinase and leads to its inactivation. Co-expression of the murine gene 3CH134 and the erk2 MAP kinase isoform in mammalian cells leads to the dephosphorylation and inactivation of the MAP kinase[11]. Furthermore, it has been shown recently that this phosphatase gene can also block cellular DNA synthesis induced by an activated version of the ras oncogene in rat embryo fibroblasts[12].

For present purposes, the terms "Mitogen-activated protein-kinase", "MAP kinase" and "MAPK" all apply to protein kinases that are activated by dual phosphorylation on threonine and tyrosine residues. This may be in response to a wide array of stimuli. Different MAP kinases are activated in response to different extracelluar stimuli, including (depending on the MAPK) stress, osmotic stress, mating pheromone (in yeast), growth factors, TNF, IL-1 and LPS. Map kinases include SMK1, HOG1,

MPK1, FUS3/KSS1, spk1, ERK1/ERK2, JNK/SAPK and p38. "MAP kinase phosphatase" activity or function is the ability to dephosphorylate one, or sometimes both, of the threonine and tyrosine residues on a MAP kinase, which residues are phosphorylated on activation of the MAP kinase. Thus, MAP kinase phosphatases are capable of hydrolysing either, or both, phosphothreonine and phosphotyrosine residues on a MAP kinase.

Martell et al, (October 1995). J. Neurochem. 65: 1823, describe the cloning of a protein tyrosine kinase abundant in brain. Theodosiou et al (1996) Hum. Molec. Genet. 5: 675 report the cloning of the murine M3/6 cDNA which is also described herein.

The present invention has arisen from the characterisation of a region in cosmids corresponding to yeast artificial chromosomes[22] during which a series of cDNA clones were identified. One of these, designated M3/6, was isolated from a mouse adult brain cDNA library. The cDNA of the invention shows homology to a family of phosphatases and appears to define a new subfamily of phosphatases. Significantly, this cDNA contains a translated complex repeat at its 3'end which may be polymorphic.

We have thus surprisingly found two genes that encode new proteins that appear to be a new members of the dual-specificity phosphatase family. We have called the new murine protein M3/6. A cDNA sequence of murine M3/6 is presented as Figure 1. Another cDNA sequence of the gene is presented as Figure 2, together with a translation of the open reading frame. All amino acid sequences represented herein are represented in the conventional N- to C- terminal direction, in the standard one letter code, unless this is specified to the contrary.

The partial cDNA sequence of the human homologue, Hb5, is shown in Figure 3. The cDNA sequence has been cloned. It shows about 81% identity at the nucleotide level to the murine sequence. Figure 4 shows an alignment between the open reading frame of the murine (top sequence) and human genes. The murine gene contains a trinucleotide repeat region which is not present in the human

hVH-5 gene. Excluding this region, the two protein sequences are about 90% homologous (where this term means amino acid identity).

The invention provides a protein of murine, human or other mammalian origin based on the cDNA information provided herein. As shown in Example 2, mRNA can be detected in the eye, brain, lung and other tissues of mice and human fetal liver, kidney, lung and brain tissue using the cDNA of the present invention, the cloning of which is described in Example 1. Translation of the cDNA obtained from Example 1 in a coupled reticulocyte assay indicated that it encodes a polypeptide of approximate molecular weight of 80 kD.

Thus the invention provides a murine phosphatase designated M3/6 or a human or other mammalian homologue thereof which phosphatase is characterized by the following features:

(a) it is encoded by a cDNA sequence obtainable from a mammalian brain cDNA library, said DNA sequence being selectively detectable with a murine DNA sequence as shown in Figure 1 or one or more of the human DNA sequences shown in Figure 3; and

(b) it comprises a phosphatase catalytic domain of the sequence VHCXAGXXRSX, where X is any amino acid.

Preferably the catalytic domain sequence is VHCLAGISRSA.

The protein desirably has the additional feature of either:

(c') tyrosine phosphatase activity, or:

(c") threonine phosphatase activity.

Preferably it has both activities.

The protein preferably has one or more of the additional features:

(d) it has a sequence of 299 amino acids at its N-terminus which are substantially as the M3/6 sequence shown in Figure 5;

(e) it has a cytoplasmic location in at least some cell types;

(f) it is encoded by an mRNA of approximately 5kb; and

(g) it comprises a C-terminal region rich in serine.

In a preferred aspect, the phosphatase is M3/6 of murine origin, in which case at least one of features (c') and (c"), together with all of features (d) to (g) may be present. However the phosphatase may also be of human origin. In this case at least one of features (c') and (c") plus feature (d) may be present.

The term "selectively detectable" means that the cDNA used as a probe is used under conditions where a target cDNA of the invention is found to hybridize to the probe at a level significantly above background. The background hybridization may occur because of other cDNAs present in the brain cDNA library. In this event background implies a level of signal generated by interaction between the probe and a non-specific cDNA member of the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target cDNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with $^{32}P$.

Suitable conditions may be found by reference to the Examples. The cDNA of Figure 1 can detect both murine and human DNA at 0.1xSSC, 0.1% SDS at 55°C (where 1xSSC is 0.15M sodium citrate, 0.15 M sodium chloride at pH7.5).

Tyrosine and threonine phosphatase activity assays are generally well knonw in the art and any suitable assay may be used. Reference may be made for example to Fischer *et al*, 1991, *Science* 253; 401-406 or Zeng and Guan, 1993, *J. Biol. Chem.*, 268;16116-16119.

Preferably all the proteins of the invention will have dual specificity phosphatase activity, namely they are capable of dephosphorylation at both Ser/Thr and Tyr residues. However, it should be borne in mind that phosphatase activity for the M3/6 protein has not yet been demonstrated, although homology with other sequences (see Example 1) strongly suggest this. It is therefore possible that this and the other polypeptides of the

invention may in fact have alternative and/or additional activities, functions or roles.

The sequence of 299 amino acids at the N-terminus of proteins of the invention will be at least 80%, preferably at least 90%, more preferably at least 95% and most preferably at least 97.5% homologous to the M3/6 sequence of Figure 5.

The cytoplasmic location of the protein of the invention may be determined in accordance with methods described in the accompanying examples. Cells which may be examined to determine cellular localization are preferably neuronal cells such as PC12 cells in at least some cell types.

The cDNA is it is encoded by an mRNA of approximately 5kb. It will be appreciated that determination of mRNA size is often (as is the present case) established by northern blotting techniques and thus is limited in accuracy by the limitations of this technique. In addition there will be some heterogeneity of the size of the polyA tail of mRNA molecules. An approximate size of 5kb will be understood by those of skill in the art to have a tolerance of at least $\pm$0.5 kb. Nonetheless, approximate mRNA size is still a useful characteristic for determining the identity of a protein.

We have found that the C-terminal region of the murine protein of encoded by the cDNA sequences of the invention are rich in serine, and also glycine. The present invention can thus be broadly thought of as relating to a phosphatase, such as a dual specificity phosphatase, that possesses at least one region of amino acid repeats at least in the murine variant of such a protein. Such repeats will be a contiguous and continuous repeat of the same amino acid. The repeat, or each repeat, may be of glycine (G) and/or serine (S) residues. Each repeat may be at least four or five, such as ten, amino acids in length. The repeat may be no longer than 20, such as 30 or 40, amino acids.

The presence of at least one stretch of 19 contiguous serine residues in the C-terminal region (i.e. within 150 or 200 amino acids) is indicative of at least the murine form of the protein of the present invention. This C-terminal region may also comprise a stretch of at least 17 contiguous glycine residues.

In addition to M3/6 and mammalian homologues thereof the present invention also contemplates:

(a) an allelic variant of such proteins;

(b) a protein at least 80% homologous to such proteins;

(c) a fragment of any one of such proteins (or (a) or (b) having phosphatase activity and being of at least 15 amino acids long; or

(d) a fusion protein comprising such proteins (or any one of (a) to (c).

All proteins and polypeptides within this definition are referred to below as proteins or polypeptide(s) according to the invention.

A protein or polypeptide of the invention will preferably be in substantially isolated form, i.e in a form in which it is free of other polypeptides with which it may be associated in its natural environment (eg the body). It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and yet still be regarded as substantially isolated.

The polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, eg. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

Mutant proteins or polypeptides are also contemplated in accordance with the invention. These will possess one or more mutations each of which is one or more additions, deletions, or substitutions of amino acid residues. Preferably the mutations will not affect, or substantially affect, the structure and/or

function and/or properties of the polypeptide. Thus, mutants may suitably possess phosphatase activity, preferably dual specificity phosphatase activity. Mutants can either be naturally occurring (that is to say, purified or isolated from a natural source) or synthetic (for example, by performing site-directed mutagenesis on the encoding DNA). It will thus be apparent that polypeptides of the invention can be either naturally occurring or, preferably, recombinant (that is to say prepared using genetic engineering techniques).

An allelic variant will be a variant which will occur naturally in a human or murine animal and which will dephosphorylate in a substantially similar manner to the proteins of the invention.

Similarly, a species homologue of the M3/6 protein will be the equivalent protein which occurs naturally in another species, eg. other than mouse or human, and which performs the equivalent or similar function in that species. Within any one species, a homologue may exist as several allelic variants, and these will all be considered homologues of the protein. Allelic variants and species homologues can be obtained by following the procedures described herein for the production of a protein of Example 3 and performing such procedures on a suitable cell source, eg from human or a rodent, carrying an allelic variant or another species. Since the protein may be evolutionarily conserved it will also be possible to use a polynucleotide of the invention to probe libraries made from human, rodent or other cells in order to obtain clones encoding the allelic or species variants. The clones can be manipulated by conventional techniques to identify a polypeptide of the invention which can then be produced by recombinant or synthetic techniques known *per se*. Preferred species homologues include mammalian or amphibian species homologues.

A protein at least 80% homologous to the M3/6 protein is included in the invention, as are proteins at least 90% and more preferably at least 95% homologous to this protein. This will generally be over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous amino

acids. Methods of measuring protein homology are well known in the art and it will be understood by those of skill in the art that in the present context. Homology is usually calculated on the basis of amino acid identity (sometimes referred to as "hard homology").

Generally, polypeptide fragments of a M3/6 protein or its allelic variants or species homologues thereof capable of exhibiting phosphatase activity will be at least 10, preferably at least 15, for example at least 20, 25, 30, 40, 50 or 60 or 100 amino acids in length.

It will be possible to determine whether the proteins or polypeptides of the invention exhibit phosphatase activity using standard routine techniques, a suitable test being given later in this specification in Example 6. Alternatively one may examine the sequence of the protein to see if it possesses the characteristic phosphatase catalytic domain, namely: (I/V)HCXAGXXR(S/T)G, wherein X represents any amino acid. In the M3/6 polypeptide of the invention this potential catalytic domain is found at 244-254 of the protein encoded by Figure 2, except in both cases the C-terminal G is replaced by A.

Preferred fragments of proteins of the invention include those which exhibit phosphatase activity and/or possess the above catalytic domain sequences. The Examples presented herein describe a number of methods to analyze the function of the protein and these may be adapted to assess whether or not a polypeptide possesses certain activities.

In Figure 5 the conceptual translation of half of (the 5' end of) the M3/6 coding sequence is aligned and compared with the human CL100 MAP kinase phosphatase protein sequence. A high degree of homology between the sequences can be seen, further indicating phosphatase activity for the M3/6 protein.

A polypeptide of the invention may be labelled with a revealing or detectable label. The (revealing) label may be any suitable label which allows the polypeptide to be detected. Suitable

labels include radioisotopes, e.g. $^{125}I$, enzymes, antibodies and linkers such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of phosphatases in a sample.

A polypeptide or labelled polypeptide according to the invention may also be fixed to a solid phase, for example the wall of an immunoassay dish.

In a second aspect of the invention, there is provided a polynucleotide which comprises:

(a) a sequence encoding a protein or polypeptide of the invention as defined above or the complement of said sequence;

(b) a sequence of nucleotides shown in Figure 1 or Figure 2;

(c) a sequence capable of selectively hybridising to a sequence in either (a) or (b); or

(d) a fragment of any of the sequences in (a) to (c).

The polynucleotide of the present invention is suitably in substantially isolated or purified form.

Polynucleotides of the invention include the DNA sequence of Figure 1 and fragments thereof capable of selectively hybridizing to the sequence of Figure 1. Polynucleotides of the invention also include polynucleotides comprising human cDNA characterized by the presence of one or more of the sequences shown in Figure 3.

The present invention in one embodiment provides a nucleic acid sequence comprising the nucleotide sequence according to Figure 1. This (murine) protein has an ATG initiation codon as shown in Figures 1 and 2. Amino acid residues encoded by the protein of Figure 2 - which are also encoded by the sequence of Figure 1 - show high homology to the cdc25 PTP of yeast[28] at residues 29-49 and 117-136. The sequence shows high homology to several PTPs in the public database EMBLGENBANK. This M3/6 gene is

murine; parts of the gene encoding for the human homologue (Hb5) are shown in Figure 3.

The gene that encodes the protein we have called M3/6 (which appears to be a tyrosine phosphatase), contains a complex triplet distal to the catalytic domain which is translated into protein. This domain comprises a run of four serine residues which is followed by a run comprising 17 glycine residues which in turn is followed by a further run comprising 23 serine residues which is interrupted near the N-terminal section by a single asparagine.

It is thought that this repeat might cause instability of this domain if it expands. Any expansion of this triplet repeat may disrupt the normal activity of the protein in the cell and lead to a disease phenotype in a way similar to other neurological disorders. This protein is highly expressed in the brain with much lower levels in liver and spleen tissues.

It will be appreciated that in polynucleotides of the invention, which encompass nucleic acid encoding a polypeptides of the invention, triplet repeats of the codons encoding the repeated amino acid residues may be present. Such codons may encode for either glycine and/or serine residues. Such triplet repeats may be at least 15, such as at least 30, bases in length, generally up to a maximum of 60 nucleotides.

If glycine residues are repeated, then triplet repeats of $(GGC)_n$ or $(GGT)_n$ (which are 2 of the 4 codons encoding Gly) can be present. Here the number n is an integer, suitably from 4, 5 or 10 up to 20. For serine, repeats of $(AGT)_m$ or $(AGC)_m$ (this residue has a degeneracy of 6) may exist. The variable m is also an integer, such as from 4 to 20.

The polynucleotide of the invention may also comprise RNA. It may also be a polynucleotide which includes within it synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothionate backbones,

addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the oligonucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or lifespan of oligonucleotides of the invention used in methods of therapy.

A polynucleotide capable of selectively hybridizing to the DNA of Figure 1, 2 or 3 will be generally at least 70%, preferably at least 80 or 90% and optimally at least 95% homologous to the DNA of Figure 1, 2 or 3 over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides. These polynucleotides are also within the invention.

A polynucleotide of the invention will be in substantially isolated form if it is in a form in which it is free of other polynucleotides with which it may be associated in its natural environment (usually the body). It will be understood that the polynucleotide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polynucleotide and it may still be regarded as substantially isolated.

A polynucleotide according to the invention may be used to produce a primer, e.g. a PCR primer, or a probe e.g. labelled with a revealing or detectable label by conventional means using radioactive or non-radioactive labels, or the polynucleotide may be cloned into a vector. Such primers, probes and other fragments of the DNA of Figure 1, 2 or 3 will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed within the invention.

Polynucleotides, such as a DNA polynucleotides according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. It may be also cloned by reference to the techniques disclosed herein.

The invention includes a double stranded polynucleotide comprising a polynucleotide according to the invention and its complement.

A third aspect of the invention relates to an (eg. expression) vector suitable for the replication and expression of a polynucleotide, in particular a DNA or RNA polynucleotide, according to the invention. The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the polynucleotide and optionally a regulator of the promoter. The vector may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. The vector may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector may also be adapted to be used *in vivo*, for example in a method of gene therapy.

Vectors of the third aspect are preferably recombinant replicable vectors. The vector may thus be used to replicate the DNA. Preferably, the DNA in the vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by a host cell. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences. Such vectors may be transformed or transfected into a suitable host cell to provide for expression of a polypeptide of the invention.

A fourth aspect of the invention thus relates to host cells transformed or transfected with the vectors of the third aspect. This may allow for the replication and expression of a polynucleotide according to the invention, such as the sequence of Figure 1, or the open reading frame thereof. The cells will

be chosen to be compatible with the vector and may for example be bacterial, yeast, insect or mammalian.

A polynucleotide according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense polynucleotides may also be produced by synthetic means. Such antisense polynucleotides may be used in a method of controlling the levels of the M3/6 phosphatase protein in a cell and/or tissue.

Thus, in a fifth aspect the invention provides a process for preparing a polypeptide according to the invention which comprises cultivating a host cell (eg. of the fourth aspect) transformed or transfected with an (expression) vector of the third aspect under conditions providing for expression (by the vector) of a coding sequence encoding the polypeptide, and recovering the expressed polypeptide.

The invention in a sixth aspect also provides (monoclonal or polyclonal) antibodies specific for a polypeptide of the invention. Antibodies of the invention include fragments thereof as well as mutants that retain the antibody's binding activity. The invention further provides a process for the production of monoclonal or polyclonal antibodies to a polypeptide of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using a polypeptide of the invention as an immunogen. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention and recovering immune serum.

In view of the presence of sequences in proteins of the present invention which are substantially homologous to sequences present in other proteins, particularly phosphatases, the antibodies will preferably be selective for the M3/6 protein and its mammalian homologues, i.e they will not recognize epitopes found on other phosphatases, particularly tyrosine/threonine phosphatases.

Fragments of monoclonal antibodies which retain antigen binding activity, such Fv, F(ab') and $F(ab_2)'$ fragments are included in this aspect of the invention. In addition, monoclonal antibodies according to the invention may be analyzed (eg. by DNA sequence analysis of the genes expressing such antibodies) and humanized antibody with complementarity determining regions of an antibody according to the invention may be made, for example in accordance with the methods disclosed in EP-A-0239400 (Winter).

The present invention further provides compositions comprising the antibody or fragment thereof of the invention together with a carrier or diluent. Polypeptides, polynucleotides, vectors and hosts of the invention can be present in compositions together with a carrier or diluent. These compositions include pharmaceutical compositions where the carrier or diluent will be pharmaceutically acceptable.

Pharmaceutically acceptable carriers or diluents include those used in formulations suitable for oral, rectal, nasal, topical (including buccal and sublingual), vaginal or parenteral (including subcutaneous, intramuscular, intravenous, intradermal, intrathecal and epidural) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Such methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In general the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

For example, formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostatis and solutes which render the formulation isotonic with the blood of the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents, and liposomes or other microparticulate systems which are

designed to target the polypeptide to blood components or one or more organs.

Polynucleotides, vectors, host cells and polypeptides, according to the invention, and antibodies or fragments thereof and compositions comprising them may be used for the treatment, regulation or diagnosis of conditions, in a mammal including man. Such conditions include those associated with aberrant (eg due to a mutation in the gene sequence) expression of one or more of the M3/6 or Hb5 proteins or related family members. Treatment or regulation of conditions with the above-mentioned moieties, especially polypeptides, antibodies, fragments thereof and compositions etc. will usually involve administering to a recipient in need of such treatment an effective amount of a polypeptide, antibody, fragment thereof or composition, as appropriate.

The present invention further provides a method of performing an immunoassay for detecting the presence or absence of a polypeptide of the invention in a sample, the method comprising:

(a) providing an antibody according to the invention;

(b) incubating the sample with the antibody under conditions that allow for the formation of an antibody-antigen complex; and

(c) detecting, if present, the antibody-antigen complex.

Vectors carrying a polynucleotide according to the invention or a nucleic acid according to the invention may be used in a method of gene therapy. Methods of gene therapy include delivering to a cell in a patient in need of treatment an effective amount of a vector capable of expressing in the cell a polypeptide of the invention.

Such vectors are preferably viral vectors. The viral vector may be any suitable vector available in the art for targeting particular cells. For example, Huber *et al* (Proc. Natl. Acac. Sci. USA (1991) 88, 8039) report the use of amphotrophic retroviruses for the transformation of hepatoma, breast, colon

or skin cells. Culver *et al* (Science (1992) 256; 1550-1552) also describe the use of retroviral vectors in virus-directed enzyme prodrug therapy, as do Ram *et al* (Cancer Research (1993) 53; 83-88). Englehardt *et al* (Nature Genetics (1993) 4; 27-34 describe the use of adenovirus based vectors in the delivery of the cystic fibrosis transmembrane conductance product (CFTR) into cells.

The invention also contemplates (diagnostic) assays. This might involve conducting an assay to find a dephosphorylation modulator, such as an inhibitor of dephosphorylation, or in other words an inhibitor of the polypeptides of the invention. It is thought that certain proteins (such as MAP kinases) are deactivated by dephosphorylation. Therefore, an inhibitor of dephosphorylation is likely to inhibit deactivation.

Thus, one assay contemplated by the invention is to identify a modulator of the phosphatase polypeptides of the invention. The assay may comprise contacting a potential chemotherapeutic agent with a protein, such as an enzyme, that will usually be dephosphorylated by a phosphatase polypeptide of the invention, and observing the phosphorylation state of the enzyme. The enzyme may be present in an extract from a cell which contains that enzyme. Enzymes contemplated include kinases, such as MAP kinases.

The polynucleotides of the invention may thus find use as probes in diagnosis, in particular diagnosis or prognosis of tumours associated with deletions in the chromosome 11, particularly 11p15 and more especially in 11p15.5. Such tumours include brain or lung tumours.

These probes may be used to detect polynucleotides of the invention, which detection may indicate that an individual has, or possesses predisposition to, a disease or disorder such as a neurodegenerative or proliferative disorder. Suitable probe detection and hybridisation techniques are well known in the art.

The M3/6 and Hb5 genes of the invention may be responsible, if mutated, for various neurodegenerative or proliferative diseases. Several repeat regions have been identified in M3/6, namely

triplet repeats that encode glycine (one repeat of 17 amino acids) and serine (3 repeats of 4, 5 and 19 amino acids respectively). The present invention thus relates to the diagnosis of susceptibility to disorders such as neurodegenerative or proliferative disorders by detecting the presence or absence of these repeat regions. By use of the unmutated gene or protein individuals may be treated that possess a neurodegenerative disease or disorder.

The probes of the present invention are hybridisable to polynucleotides of the invention suitably under low stringency conditions. However, it is preferred that hybridisation take place under high stringency conditions. By low stringency conditions one envisages 3X SSC (0.5M sodium chloride, pH7.5) at room temperature. High stringency conditions that are envisaged are 0.1X SSC (0.1M sodium chloride, pH7.5) at 65°C.

It will be apparent that probes contemplated may be capable of hybridising to the region of triplet repeats. In the M3/6 gene, this is encompassed by nucleotides 1756 to 1875 of the sequence shown in Figure 1. Such probes will be at least 15, preferably at least 20, for example 25, 30 or 40, nucleotides in length.

The invention can thus provide a method of screening for susceptibility to a disease or disorder such as a neurodegenerative or proliferative disorder, which method comprises detecting, and possibly analysing, the triplet repeat region (present in polynucleotides of the invention) of an individual.

The method may involve the polymerase chain reaction (PCR). It is preferred that such methods will not require the use or radiolabelled nucleotides. Detection of a normal triplet repeat, such as is present in the M3/6 protein, may indicate than an individual's susceptibility to the neurodegerative disorder or disease is low.

The method may also extend to diagnosing susceptibility to a disorder or disease such as a neurodegenerative or proliferative

disorder which method comprises detecting, if present, an amplification in a GGC, GGT, AGT or AGC repeat in a region of the human or animal genome that corresponds to the location of a polynucleotide of the present invention.

An amplification in the polynucleotide repeat may be determined by removing a sample or genomic DNA from the patient, carrying out a PCR with primers upstream and downstream of the repeat region, and determining the amount of nucleic acid produced. PCR generally does not occur to a substantial extend across genomic DNA comprising a repeat of 30 repeats or more. Substantial amounts of nucleic acid are only produced by PCR carried out on a DNA fragment in which there is little or no amplification of the nucleotide repeat, i.e. less than 30 repeats.

In the accompanying drawings, which are provided to illustrate the present invention:

Figure 1 gives a cDNA sequence of the M3/6 gene;

Figure 2 gives another cDNA sequence of the M3/6 gene and the open reading frame thereof;

Figure 3 shows portions of the cDNA sequence encoding the Hb5 human homologue;

Figure 4 shows an alignment between the open reading frame of the murine (top sequence) M3/6 and human Hb5 genes. The latter is as disclosed in Martell et al, ibid;

Figure 5 gives the N-terminal sequence of the M3/6 protein, and aligns it with the CL100 phosphatase, from which two proteins a consensus sequence is derived; and

Figure 6 is a graphical representation of the hVH-5 gene structure.

The following Examples describe the isolation and characterization of the novel protein and DNA of the invention from murine and human sources. However, other e.g. mammalian sources are within the scope of the present invention and other mammalian homologues of the protein may be isolated in an analogous manner. The Examples are presented here by way of illustration and are not to be construed as limiting on the invention.

EXAMPLE 1 - Sequence data

A novel nucleic acid sequence (murine M3/6) is presented which encodes a putative dual specificity threonine-tyrosine phosphatase which may be used in the characterisation of signalling mechanisms in brain and muscle. The presence of a complex trinucleotide repeat, located at the 3'end of this sequence and which is translated, makes this phosphatase gene a candidate for a human disease caused by repeat expansion or mutation. Fragments of the human gene homologue (Hb5) are also presented.

**Isolation of M3/6.**

A human fragment from a yeast artificial chromosome (YAC) was isolated. Such YACs contain well over 50kb and to produce smaller, manageable sized segments for analysis were subcloned into cosmids of 45kb or less each. A series of cDNA clones were identified from these cosmids.

M3/6 was isolated from a mouse brain cDNA library constructed from oligo dT and random primed cDNA (Blake, D.J., Nawrotzki, R. and Davies, K. E. Isoform diversity of the murine 87K postsynaptic protein; submitted), cloned into the EcoRI site of the vector pcDNAII.

pcDNAII is a 2.9 kb plasmid vector from Invitrogen and contains the Ampicillin resistance gene. The M3/6 cDNA was isolated from the host cells XL1-Blue, by standard alkaline lysis method of preparing plasmid DNA. The 2.5 kb insert containing the entire M3/6 cDNA was released from the vector by digestion with the restriction enzyme EcoRI. The insert was separated from the vector by gel electrophoresis on 1% agarose and purified using spin columns. The purified insert was radiolabelled using Amersham megaprime labelling kit. M3/6/4e is a deletion derivative of M3/6 generated by the Erase-a-Base system. It encompasses nucleotides 1 to 1000 of M3/6 and can be released from the vector using the restriction enzymes EcoRI and XbaI.

**Sequencing of M3/6.**

A nested set of deletion clones of the 2.5 kb cDNA was generated using the Erase-a-Base System commercially available from Promega. These clones were sequenced using double stranded sequencing protocol from USB. Sequencing reactions were resolved on a standard 6% acrylamide gel and visualised by autoradiography after overnight exposure at room temperature. Sequence analysis was done using the GCG Wisconsin package version 8.

Sequence comparisons (see Figure 5) suggest that the M3/6 novel gene described is also a dual specificity phosphatase and will be able to dephosphorylate MAP kinase. In addition, portions of the murine M3/6 gene show considerable homology to the human Hb5 gene homologue.

The human Hb5 gene was isolated by screening a Clontech (commercially available) human foetal brain cDNA library with the M3/6 sequence.

EXAMPLE 2 - Protein distribution in tissues

**RNA extraction and Northern blotting.**

RNA was extracted from mouse tissue following the method of Chomczynski, P. and Sacchi, N. (1987, Anal. Biochem. 162,156-159). poly $A^+$ plus RNA was prepared from 100$\mu$g of total RNA using the Dynabeads mRNA purification kit from Dynal. Northern blots were prepared according to Current protocols in Molecular Biology. The human fetal tissue Northern was obtained from Clontech. Hybridisation was carried out at 42°C and the blots were washed to a stringency of 0.1xSSC, 0.1%SDS at 55°C. The blots were visualised by autoradiography after exposure for one-two days at -70°C.

The results of the hybridisation of the M3/6 clone to Northern blot containing several mouse tissues were examined. A band at 5kb in mouse eye and brain was seen, but no bands of significance were seen in spleen, skeletal muscle, small intestine or liver. The filter was washed at 0.1 x SSC at 50°C. A 1.8kb band is seen in mouse lung with a faint band at 5kb. A similar band at 5kb

is seen when a subclone of M3/6, designated M3/6/4e (nucleotides 1 to 1004) was used. In this case the 5kb band in lung is much stronger.

Hybridisation of M3/6/4e to a Northern blot of human fetal tissues again showed the 5kb transcript predominantly in the brain and to a lesser extent the lung, but not the kidney or liver to any significant extent. This blot is evidence of the sequence conservation of this gene between mouse and man.

## EXAMPLE 3 - Assay

***In vitro* transcription-translation assay.**

The pcDNAII vector utilises SP6 and T7 promoters which can be directly used for *in vitro* transcription-translation assays. One μg of RNase free circular plasmid DNA containing the insert was used for each reaction. The assay was performed according to the instructions provided with the Promega TNT Coupled Reticulocyte Lysate Systems. The synthesized proteins were analyzed by SDS gel electrophoresis on a 10% acrylamide gel and visualised by autoradiography after 1 hr exposure at room temperature. Luciferase-encoding plasmids were used as controls for this assay.

An analysis of the M3/6 clone in the transcription/translation coupled reticulocyte assay indicated that the protein product was 80kD indicating that the translation of the mRNA must extend through and beyond the triplet repeats. The assay was carried out using a kit from Promega according to the manufacturer's instructions.

## EXAMPLE 4 - B8 Homology

**Hybridisation of oligonucleotide M3/6-c.**

M3/6-c is a 19-mer oligonucleotide the sequence of which is CTTGGTCATCGACAGCCGG and is from the cdc25 homology region of M3/6. The oligonucleotide was radiolabelled using Promega

polynucleotide kinase and γdATP according to the manufacturer's instructions. Hybridisation was effected at 42°C and washes were done in 3xSSC, 0.1%SDS at room temperature for the cosmids and at 55°C for the cosmid subclones.

The subclone filters were washed at 3xSSC at 55°C. A strong signal was obtained. This suggests that a human sequence with high homology to this motif is present in B8. B8 contains markers (e.g.G1 and CMS1) which are in linkage disequilibrium with autosomal spinal muscular atrophy. Thus this PTP is a candidate for this motor neuron disorder and parts of it may be useful diagnostically.

## EXAMPLE 5 - M3/6 Expression

Proof that M3/6 encodes a cytoplasmic protein, which was expressed in PC12 cells, was derived from the following experiment. A plasmid capable of directing expression of the M3/6 gene in a mammalian cell line was constructed by cloning the M3/6 cDNA into the polylinker of the vector pEFmycpLINK[29]. This results in the expression of a fusion protein between the myc epitope (MEQKLISEEDL), recognised by the monoclonal antibody 9E10, and the protein of the invention under the control of the Elongation Factor gene promoter. This DNA construct was microinjected into PC12 cells. These cells are able to undergo neurite outgrowth typical of neuronal cells when stimulated with Nerve Growth Factor (NGF). Expression of M3/6 was monitored by staining with the α-myc epitope antibody 9E10. This revealed the surprising and novel finding that M3/6 encodes a cytoplasmic protein. Where the localization of other potential MAP kinase phosphatases has been determined, this has been exclusively nuclear. Furthermore expression of M3/6 failed to block NGF-stimulated neurite outgrowth which is surprising as expression of a MAP kinase phosphatase might be expected to block this MAP kinase dependent process.

Given the unusual cytoplasmic location of M3/6 it is possible that mutation or frameshift at the triplet repeat could lead to a change in the subcellular location of the protein. This might

lead to its relocation to the nucleus which may be a 'default' location for proteins of this family. Potentially this could lead to a loss or gain of function.

EXAMPLE 6 - Tests for MAP kinase phosphatase activity

The putative phosphatase was expressed and purified from bacterial or insect or mammalian cells followed by incubation with *in vitro* $^{32}P$ phosphorylated MAP kinase. Dephosphorylation of the MAP kinase can be assayed by gel electrophoresis followed by autoradiography.

An alternative assay involves co-expression of the putative phosphatase and a myc-epitope tagged version of MAP kinase in COS cells. Stimulation of these transfected cells with e.g. serum or EGF leads to a mobility shift in the MAP kinase which is revealed by gel electrophoresis, western blotting and probing with a myc epitope recognising antibody. Co-expression of a MAP kinase phosphatase should lead to the abolition of this mobility shift.

This specification describes the identification of a novel gene encoding a novel protein that is highly likely to be a phosphatase which is a member of a sub-family of dual specificity threonine-tyrosine phoshatases expressed in neuronal tissue. It has a motif which shows very high homology to the yeast cdc25 yeast tyrosine phophatases and possesses the characteristic conserved catalytic domain of all phosphatases. A transcription/translation coupling experiment (Example 3) has confirmed the presence of an expressible open reading frame and strongly suggests that the complex repeat is expressed as part of the 3'domain of the molecule. Since this may expand by replication slippage or other mechanisms as in other neurological disorders, any change in the size of this triplet repeat may give rise to molecular pathology.

The presence of the crosshybridisation of this sequence to human sequences derived from the candidate gene region for SMA makes this a candidate gene for the disorder. Since the gene is

expressed in brain and lung, but predominantly in brain, any change in the function of the gene might give rise to a neurological disorder. This can be tested by using the sequence or part of the sequence as a probe to hybridise to DNA from patients with such diseases. Alterations in DNA from patients might also be seen using PCR primers derived from the corresponding human sequence. A change in the protein might also be detected using antibodies raised from peptides or expressed portions of this sequence and the investigation of muscle biopsies.

## EXAMPLE 7

## Chromosomal Localization and Genomic Organization of Hb5.

HB5 cDNA was used previously for fluorescence in situ hybridisation (FISH) analysis on human metaphase chromosomes (Theodosiou et al., 1996). It mapped to three locations with the principal peaks being on 10q11.2 and distal 11p15, with a further peak on 10q22. To further refine the chromosomal localisation of HB5 both a human chromosome 11 cosmid library (Smith et al., 1993) and a total human genomic PAC library (Ioannou et al., 1994) were screened with M3/6-4e and HB5 respectively. Four cosmids and nine PACs were isolated. All nine PACs gave an identical PstI restriction pattern when probed with HB5 that was entirely different from that of the cosmids, whose pattern was similarly identical. However both PACs and cosmids showed cognate bands with PstI-digested total human genomic DNA. Thus cDNA-positive cosmid bands of approximately 2.3kb (doublet), 1.6kb (dimorphic with a 2.0kb band) and 1.05kb and PAC bands of 1kb and 3kb are seen in digests of total human DNA. The possibility of other copies of this gene or related genes is suggested by other bands seen in the genomic DNA digests at approximately 1kb and 4kb. To assign a chromosomal localisation to these two separate genomic clone contigs one cosmid (cSRL 15a6) and two of the PACs (86N13 and 234B10)were used in FISH experiments. The cosmid maps uniquely to 11p15.5, whilst the PACs both map to 10q11.2. No signal from either was seen corresponding to the minor peak identified by FISH with HB5 cDNA on 10q22 (Theodosiou et al., 1996).

PCR analysis of the PAC clones using oligonucleotide primers based on the cDNA sequence showed that in all cases the size of the PCR amplification product was exactly the same size in both PACs and cDNA. This suggests that the PAC, and hence the chromosome 10q11.2 copy of the gene is intronless and presumably a pseudogene. In contrast, the PCR products from the cosmid were, for the most part larger than the cDNA, suggesting the presence of introns. Subclones of the cosmid that were positive by hybridisation to the cDNA were sequenced to determine the intron/exon boundaries and the flanking intronic sequences by comparison with the cDNA sequence. A graphical representation of the hVH-5 gene is shown in Fig. 6. The 1875bp open reading frame coding for hVH-5 is distributed over 6 exons, the smallest of which is 124bp (Table 1). The sizes of exon 1 and exon 5 have not been determined since neither the transcription start site nor the polyadenylation signal for this gene have been found but, given a mRNA size of 5kb (Martell et al., 1995; Theodosiou et al., 1996) and assuming no 5' or 3' untranslated exons (as in CL100(Kwak et al., 1994)the gene is spread over not more than 13kb.

The introns range in size from 193bp to approximately 4.75kb, with the second intron being by far the largest. The first exon contains the initiating methionine and the first CH2 (cdc25 homology 2) domain. The second CH2 domain is split between exons 2 and 3. Exon 5 contains the entire conserved catalytic PTPase domain, whilst the entire PEST (proline, glutamic acid, serine and threonine-rich) domain is contained within exon 6, which also contains the translation termination codon and all the 3' UTR so far described.

Using conserved primers flanking the trinucleotide repeat found in the mouse cDNA, M3/6 (Theodosiou et al., 1996) PCR analysis using the cosmids and PACs as templates showed that both the chromosome 10 and 11 copies of the human gene gave the size of product predicted from the human hVH-5 cDNA sequence. This was confirmed in the chromosome 11 cosmid by sequencing using these same primers. In addition, no polymorphism for this repeat was noted among a small number of human individuals. Thus no

evidence was found for a copy of this gene containing the complex trinucleotide repeat found in mouse.

EXAMPLE 8 - Loss of Heterozygosity of hVH-5 in lung tumours

The region of chromosome 11q to which hVH-5 maps has been implicated in the development of non-small cell lung cancer (NSCLC) breast cancer, rhabdomyosarcoma, Wilm's tumour , bladder cancer and testicular cancer (see Bepler and Garcia, 1994). Given the previously suggested potential tumour-supressor activity of hVH-5 (Martell et al., 1995; Theodosiou et al., 1996) we investigated loss of heterozygosity at this locus in 15 lung tumour samples. Eight of these were heterozygous for a PstI polymorphism in DNA from normal blood, one of which showed loss of heterozygosity in DNA from the corresponding tumour.

EXAMPLE 9 - Analysis of Methylation at the hVH-5 locus

A number of genes which map to human chromosome 11p15.5 are imprinted, that is only one of the parental alleles is expressed in somatic cells (Barlow, 1995) These include IGF2 and H19 (Rainier et al., 1993). One phenomenon associated with imprinting is the differential methylation of the parental alleles (Barlow, 1995) . It has recently been suggested that imprinted genes have few and small introns (Hurst et al., 1996) Since hVH-5 both maps to chromosome 11p15.5 and has few, small introns, there is the possibility that this gene might also be imprinted. Imprinting of the gene from normal adult brain and lung is studied as is imprinting in fetal and tumour cells. Comparison of the patterns of imprinting may be used to provide diagnostic and/or prognostic assays of disease status. Diseases include proliferative diseases of lung and/or brain tissue.

**REFERENCES**

1. Mauro, L.J. and Dixon, J.E. (1994) Trends in Biochem 19: 151-155.

2. Stone. R.L. and Dixon, J.E. (1994) J Biol Chem 269: 31323-31326.

3. Hunter, T. (1995) Cell 80: 225-236.

4. Noguchi, T. et al (1993) Mol Cell Biol. 13: 5195-5205.

5. Davis, R.J. (1994) Trends in Biochem Sci 19: 470-473.

6. Marshall, C.J. (1994) Curr Opin in Genet and Dev 4: 82-89.

7. Keyse, G.M. & Emslie, E.A. (1992) Nature 359: 644-647.

8. Charles, C.H. et al (1992) Oncogene 7: 187-190.

9. Alessi, D.R. et al (1993) Oncogene 8: 2015-2020.

10. Charles, C.H. et al (1993) Proc Natl Acad Sci 90:5292-5296.

11. Sun, H. et al (1993) Cell 75: 487-493.

12. Sun, H. et al (1994) Science 266: 285-288.

13. Dubowitz,V. (1978). Muscle Disorders in Childhood, W.B. Saunders Co. Ltd., East Sussex, pp 146-190.

14. Brzustowicz, L.M. et al (1990) Nature, 344: 540-541.

15. Gilliam, T.C. et al (1990) Nature, 345: 823-825.

16. Melki ,J. et al (1990a) Lancet, 336: 271-273.

17. Melki, J. et al (1990b) Nature 344: 767-768.

18. McLean, M.D. et al (1994) Hum Mol Genet 3: 1951-1956.

19. Daniels, R.J. et al (1995) J Med Genet 32: 93-96.

20. DiDonato et al (1995) Am J Hum Genet 55: 1218-1297

21. Melki, J. et al (1994) Science, 264: 1474-1477.

22. Francis, M.J. et al (1993) Hum Mol Genet 2: 1161-1167.

23. Kleyn, P.W. et al (1993) Proc Natl Acad Sci, USA 90: 6801-6805.

24. Theodosiou, A.M. et al (1994) Am J Hum Genet 55: 1209-1217.

25. Carpten, J.D. et al (1994) Genomics 29: 351-356.

26. Roy, N. et al (1995) Cell 80: 167-178.

27. Lefebvre, S. et al (1995) Cell 80: 155-165.

28. Keyse, S.M. & Ginsburg, M. (1993) Trends in Biochem Sci 18: 377-378.

29. Marais, R. et al The EMBO Journal, 14: 3136-3145 (1995).

30. Barlow, D. P. (1995). Gametic imprinting in mammals. Science 270: 1610.

31. Bepler, G., and Garcia, B. M. (1994). Three tumor-suppressor regions on chromosome 11p identified by high-resolution deletion mapping in human non-small-cell lung cancer. Proc. Natl. Acad. Sci. U S A 91: 5513.

32. Hurst, L. D., McVean, G., and Moore, T. (1996). Imprinted genes have few and small introns. Nature Genet. 12: 234.

33. Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., and de, J. P. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genet. 6: 84.

34. Kwak, S. P., Hakes, D. J., Martell, K. J., and Dixon, J. E. (1994). Isolation and characterization of a human dual specificity protein-tyrosine phosphatase gene. J. Biol. Chem. 269: 3596.

35. Martell, K. J., Seasholtz, A. F., Kwak, S. P., Clemens, K. K., and Dixon, J. E. (1995). hVH-5: a protein tyrosine phosphatase abundant in brain that inactivates mitogen-activated protein kinase. J. Neurochem. 65: 1823.

36. Rainier, S., Johnson, L. A., Dobry, C. J., Ping, A. J., Grundy, P. E., and Feinberg, A. P. (1993). Relaxation of imprinted genes in human cancer. Nature 362: 747.

37. Smith, M. W., Clark, S. P., Hutchinson, J. S., Wei, Y. H., Churukian, A. C., Daniels, L. B., Diggle, K. L., Gen, M. W., Romo, A. J., Lin, Y., and et, a. l. (1993). A

sequence-tagged site map of human chromosome 11. Genomics 17: 699.

38. Theodosiou, A. M., Rodrigues, N. R., Nesbit, M. A., Ambrose, H. J., Paterson, H., McLellan-Arnold, E., Boyd, Y., Leversha, M. A., Owen, N., Blake, D. J., Ashworth, A., and Davies, K. E. (1996). A member of the Map kinase phosphatase gene family in mouse containing a complex trinucleotide repeat in the coding region. Hum. Molec. Genet. 5: 675.

CLAIMS

1. A polypeptide comprising a murine phosphatase designated M3/6 or a human or other mammalian homologue thereof which phosphatase is characterized by the following features:

(a) it is encoded by a cDNA sequence obtainable from a mammalian brain cDNA library, said DNA sequence being selectively detectable with a murine DNA sequence as shown in Figure 1 or one or more of the human DNA sequences shown in Figure 3; and

(b) it comprises a phosphatase catalytic domain of the sequence VHCXAGXXRSX, where X is any amino acid.

2. A polypeptide comprising:

(a) an allelic variant of a protein as defined in claim 1;

(b) a protein at least 80% homologous to a protein as defined in claim 1;

(c) a fragment of a protein as defined in claim 1 or (a) or (b) above having phosphatase activity and being of at least 15 amino acids long; or

(d) a fusion protein comprising a protein as defined in claim 1 or any one of (a) to (c) above.

3. A polypeptide according to claim 1 or 2 carrying a revealing or detectable label.

4. A polypeptide according to claim 1, 2 or 3 fixed to a solid phase.

5. A polynucleotide which comprises:

(a) a sequence encoding a protein or polypeptide as defined in claim 1 or 2 or the complement of said sequence;

(b) a sequence of nucleotides shown in Figure 1 or Figure 2;

(c) a sequence capable of selectively hybridising to a sequence in either (a) or (b); or

(d) a fragment of any of the sequences in (a) to (c).

6. A polynucleotide according to claim 5 which is a DNA polynucleotide.

7. A polynucleotide according to claim 5 or 6 which comprises at least 20 nucleotides.

8. A polynucleotide according to any of claims 5 to 7 which comprises the cDNA sequence shown in Figure 1.

9. A polynucleotide according to any of claims 5 to 8 carrying a revealing or detectable label.

10. A vector comprising a polynucleotide according to any of claims 5 to 9.

11. A vector according to claim 10 which is a recombinant replicable vector comprising a coding sequence which encodes a polypeptide as defined in claim 1 or 2.

12. A host cell comprising a vector according to claim 11.

13. A host cell according to claim 12 transformed by, or transfected with, a recombinant vector according to claim 11.

14. A host cell transformed by a recombinant vector according to claim 11 wherein the coding sequence is operably linked to a control sequence capable of providing for the expression of the coding sequence by the host cell.

15. A process for preparing a polypeptide as defined in claim 1 or 2, the process comprising cultivating a host cell according to any of claims 12 to 14 under conditions providing for expression of the recombinant vector of the coding sequence, and recovering the expressed polypeptide.

16. An antibody or a fragment thereof capable of binding to a polypeptide as defined in claim 1 or 2.

17. A screening assay for identifying a putative chemotherapeutic agent for the treatment of disease, the assay comprising:

(A) bringing into contact:

(i) a phosphorylated polypeptide which can be dephosphorylated by M3/6;

(ii) a polypeptide as defined in claim 1 or 2; and

(iii) a putative chemotherapeutic agent;

under conditions in which component (ii) would dephosphorylate component (i) in the absence of (iii); and

(B) measuring the extent to which component (iii) is able to disrupt, interfere with or inhibit dephosphorylation.

18. An assay according to claim 17 wherein the putative chemotherapeutic agent is a fragment of 10 or more amino acids of a polypeptide as defined in claim 2.

19. A method of diagnosing susceptibility to a disease or disorder, the method comprising detecting an amplification or mutation in a $(GGC)_n$, $(GGT)_n$, $(AGT)_m$ or $(AGC)_m$ repeat where n and m independently represent an integer from 2 to 20 in a region of the human genome corresponding to the location of a polynucleotide according to claim 5.

20. A method according to claim 19 wherein n is an integer from 15 to 20 and m is an integer from 4 to 20.

21. An isolated polypeptide which comprises the M3/6 sequence shown in Figure 2.

22. An isolated polynucleotide encoding the polypeptide of claim 21.

23. An isolated polynucleotide according to claim 22 which has the sequence depicted in Figure 2.

FIGURE 1

GCCAGGTCTGGCACCATGCACTAGGATACCCAGAACGCTGCAAGGCCACGCCCTCCTCAC

TTCAGGGGTCACTCTCCCCATTGCCCACCACCCCACCATGGCTGGGGATCGGCTCCCGAG

GAAGGTGATGGACGCAAAGAAACTGGCCAGCCTGCTGCGTGGCGGGCCTGGGGGACCCTT

GGTCATCGACAGCCGGTCCTTCGTGGAGTATAACAGCTGCCACGTGCTGAGCTCTGTGAA

TATCTGCTGTTCAAAGCTGGTGAAGCGGCGCCTTCAGCAGGGAAAAGTGACAATTGCTGA

GCTTATCCAGCCTGCTACACGGAGCCAGGTGGATGCCACAGAACCACAGGATGTAGTGGT

GTATGACCAGAGCACACGAGATGCCAGCGTGCTGGCAGCAGACAGCTTCCTGTCCATCCT

GCTCAGCAAGCTGGACGGCTGCTTCGACAGTGTGGCCATCCTCACAGGACGTTCGCCACC

TTCTCCTCCTGCTTCCCTGGCCTCTGTGAGGCAAGCCTGCCACTCTACCGTCCATGAGCC

TCTCTCAGCCCTGCCTGCCTGTGCCCAGTGTTGGCCTGACCCGAATCCTGCCTCACCTCT

ACCTGGGCTCTCAGAAAGATGTCTTGAACAAGGATCTGATGACCCAAAACGGAATAAGCT

ATGTCCTCAATGCCAGCAACTCCTGCCCTAAACCGGACTTCATCTGTGAGAGCCGTTTCA

TGCGTATCCCCATCAATGACAACTACTGTGAAAAGCTGCTGCCCTGGCTGGACAAGTCCA

TCGAGTTTATTGATAAAGCCAAGCTGTCCAGCTGCCAAGTCATTGTTCACTGTCTGGCTG

GCATCTCTCGCTCTGCCACCATTGCCATCGCGTACATCATGAAAACCATGGGCATGTCTT

CTGACGACGCATACAGGTTTGTGAAGGATCGGCGCCCCTCCATCTCGCCCAACTTCAACT

TCCTGGGCCAGTTGCTGGAGTATGAGAGGAGTCTGAAGCTGCTGGCTGCCCTGCAGACTG

ATGGACCTCACTTGGGGACCCCTGAGCCCCTCATGGGCCCGGCAGCAGGCATCCCACTGC

CCCGGCTGCCACCATCTACCTCAGAGAGCGCTGCCACTGGGAGCGAGGCAGCCACCGCAG

CCAGGGAGGGCAGCCCAAGTGCTGGAGGG.ATGC..TCCGATCCCCAGCACAGCTCCAGC

CACCAGCG..CGGCTGCAG.CAGG.CCTGC.GTGGCCTGCACCTCTCCTCTGACCGCCTC

CAGGACACCAACCGCCTCAAG.CGTTCCTTTTCCCTGGACATCAAGTCGGCCTATGCACC

CAGCAGGAGGCCCGACTTTCCCGGCCCNACCNGACCCCCGGTGAAGCCCCAAGCTCTNAA

GCTGACAGCCNGTCTGGGGGNACACTGGGCCTGCCCTCGCCCAGCCCAGACAGCCCGGAC

TCCGTTCCAGAGTGCCGCCCACGACC.CCGCCG.CGACGCCCCCCCGGCTAGTTCGCCTG

CCCGCTCCCCCGCGCATGGTCTGGGCCTGAACTTTGGAGACACGGCCCGGCAGACTCCAC

GGCASSCTCTCGGCCCTGTCGGCGCCCGGGCTGCCTGGCCCTGCCAGCCGGCTGGNCCCG

GGGGCTGGGTGCCGCCACTGGACTCCCCAGGCACACCGTCGCCNCCAGGNGNGCAGGGTC

CAGGCGCTGTGTTCTCCCCTTTGGCCGGGTAAGTGCAGGCGNANCTGGACCCGGTAACAG

CAGCAGCAGYGGTGGTGGTGGTGGTGGYGGCGGCGGCGGCGGCGGCGGCGGCGG

CAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG

TAGTAGTAGTAGTAGTGACCTGCGGAGGCGGGATGTGCGGACCGGCTGGCCCGAGGAGCC

TGCTGCAGATGCACAGTTCAAGAGGCGCAGCTGCCAGATGGAGTTCGAAGAGGGCATGGT

GGAGGGGCGGCACGTGCAGAGCTCCTGGCAGNCCTGGCAAGCAAACCAGCTTCTCTGGCA

GCGTGGAGGTCATCGAAGTATCGTGCCTTCAGAAGTCCCTGTGCCCTTGCTCCAGCCAGG

CCAGGTATAAATATATATTATATATAAAACACACAGAAAAGGTAAATGGTTTTAC..TGC

AATTTTTATCAAGAAGTAAATATT.CGATTTTT.ATTTATTTAAGCTAGTGATCTGGCAA

CTGTGCGGGGCGGCCCTAAAGCTCTGTTTTTACTGTCTGGTATTTAAACTGAAACAGGTT

TCTAAGCAATATGAGGCCACCTTCAATCCCAAACTGGGTTGACAGGCCTGGGCCCCTCCT

T.GCCCCTCCCCTCTGGAAACATTACTGACCTTTCAAAGAGCTGCCCAGCTTTCCTGCAC

TTTTTACATAAGAAAAAAGGGGGGGGGRAA

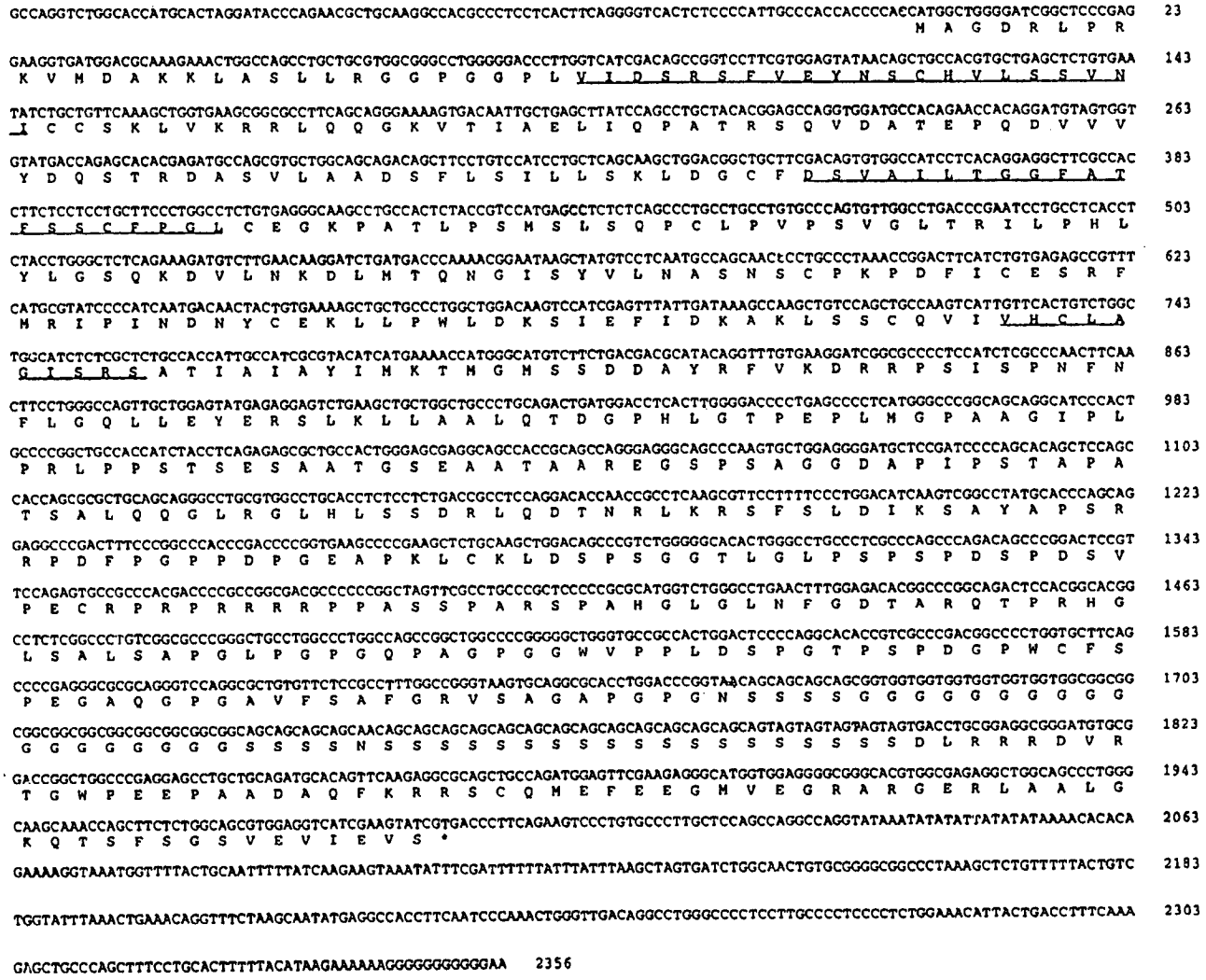

GCCAGGTCTGGCACCATGCACTAGGATACCCAGAACGCTGCAAGGCCACGCCCTCCTCACTTCAGGGGTCACTCTCCCCATTGCCCACCACCCCAECATGGCTGGGGATCGGCTCCCGAG 23
M A G D R L P R
GAAGGTGATGGACGCAAAGAAACTGGCCAGCCTGCTGCGTGGCGGGCCTGGGGGACCCTTGGTCATCGACAGCCGGTCCTTCGTGGAGTATAACAGCTGCCACGTGCTGAGCTCTGTGAA 143
K V M D A K K L A S L L R G G P G G P L V I D S R S F V E Y N S C H V L S S V N
TATCTGCTGTTCAAAGCTGGTGAAGCGGCGCCTTCAGCAGGGAAAAGTGACAATTGCTGAGCTTATCCAGCCTGCTACACGGAGCCAGGTGGATGCCACAGAACCACAGGATGTAGTGGT 263
I C C S K L V K R R L Q Q G K V T I A E L I Q P A T R S Q V D A T E P Q D V V V
GTATGACCAGAGCACACGAGATGCCAGCGTGCTGGCAGCAGACAGCTTCCTGTCCATCCTGCTCAGCAAGCTGGACGGCTGCTTCGACAGTGTGGCCATCCTCACAGGAGGCTTCGCCAC 383
Y D Q S T R D A S V L A A D S F L S I L L S K L D G C F D S V A I L T G G F A T
CTTCTCCTCCTGCTTCCCTGGCCTCTGTGAGGGCAAGCCTGCCACTCTACCGTCCATGAGCCTCTCTCAGCCCTGCCTGCCTGTGCCCAGTGTTGGCCTGACCCGAATCCTGCCTCACCT 503
F S S C F P G L C E G K P A T L P S M S L S Q P C L P V P S V G L T R I L P H L
CTACCTGGGCTCTCAGAAAGATGTCTTGAACAAGGATCTGATGACCCAAAACGGAATAAGCTATGTCCTCAATGCCAGCAACtCCTGCCCTAAACCGGACTTCATCTGTGAGAGCCGTTT 623
Y L G S Q K D V L N K D L M T Q N G I S Y V L N A S N S C P K P D F I C E S R F
CATGCGTATCCCCATCAATGACAACTACTGTGAAAAGCTGCTGCCCTGGCTGGACAAGTCCATCGAGTTTATTGATAAAGCCAAGCTGTCCAGCTGCCAAGTCATTGTTCACTGTCTGGC 743
M R I P I N D N Y C E K L L P W L D K S I E F I D K A K L S S C Q V I V H C L A
TGGCATCTCTCGCTCTGCCACCATTGCCATCGCGTACATCATGAAAACCATGGGCATGTCTTCTGACGACGCATACAGGTTTGTGAAGGATCGGCGCCCCTCCATCTCGCCCAACTTCAA 863
G I S R S A T I A I A Y I M K T M G M S S D D A Y R F V K D R R P S I S P N F N
CTTCCTGGGCCAGTTGCTGGAGTATGAGAGGAGTCTGAAGCTGCTGGCTGCCCTGCAGACTGATGGACCTCACTTGGGGACCCCTGAGCCCCTCATGGGCCCGGCAGCAGGCATCCCACT 983
F L G Q L L E Y E R S L K L L A A L Q T D G P H L G T P E P L M G P A A G I P L
GCCCCGGCTGCCACCATCTACCTCAGAGAGCGCTGCCACTGGGAGCGAGGCAGCCACCGCAGCCAGGGAGGGCAGCCCAAGTGCTGGAGGGGATGCTCCGATCCCCAGCACAGCTCCAGC 1103
P R L P P S T S E S A A T G S E A A T A A R E G S P S A G G D A P I P S T A P A
CACCAGCGCGCTGCAGCAGGGCCTGCGTGGCCTGCACCTCTCCTCTGACCGCCTCCAGGACACCAACCGCCTCAAGCGTTCCTTTTCCCTGGACATCAAGTCGGCCTATGCACCCAGCAG 1223
T S A L Q Q G L R G L H L S S D R L Q D T N R L K R S F S L D I K S A Y A P S R
GAGGCCCGACTTTCCCGGCCCACCCGACCCCGGTGAAGCCCCGAAGCTCTGCAAGCTGGACAGCCCGTCTGGGGGCACACTGGGCCTGCCCTCGCCCAGCCCAGACAGCCCGGACTCCGT 1343
R P D F P G P P D P G E A P K L C K L D S P S G G T L G L P S P S P D S P D S V
TCCAGAGTGCCGCCCACGACCCCGCCGGCGACGCCCCCCGGCTAGTTCGCCTGCCCGCTCCCCCGCGCATGGTCTGGGCCTGAACTTTGGAGACACGGCCCGGCAGACTCCACGGCACGG 1463
P E C R P R P R R R R P P A S S P A R S P A H G L G L N F G D T A R Q T P R H G
CCTCTCGGCCCTGTCGGCGCCCGGGCTGCCTGGCCCTGGCCAGCCGGCTGGCCCCGGGGGCTGGGTGCCGCCACTGGACTCCCCAGGCACACCGTCGCCCGACGGCCCCTGGTGCTTCAG 1583
L S A L S A P G L P G P G Q P A G P G G W V P P L D S P G T P S P D G P W C F S
CCCCGAGGGCGCGCAGGGTCCAGGCGCTGTGTTCTCCGCCTTTGGCCGGGTAAGTGCAGGCGCACCTGGACCCGGTAACAGCAGCAGCAGCGGTGGTGGTGGTGGTGGTGGCGGCGG 1703
P E G A Q G P G A V F S A F G R V S A G A P G P G N S S S S G G G G G G G G G G
CGGCGGCGGCGGCGGCGGCGGCGGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAGTAGTAGTAGTAGTGACCTGCGGAGGCGGGATGTGCG 1823
G G G G G G G G G S S S S N S S S S S S S S S S S S S S S S S S D L R R R D V R
GACCGGCTGGCCCGAGGAGCCTGCTGCAGATGCACAGTTCAAGAGGCGCAGCTGCCAGATGGAGTTCGAAGAGGGCATGGTGGAGGGGCGGGCACGTGGCGAGAGGCTGGCAGCCCTGGG 1943
T G W P E E P A A D A Q F K R R S C Q M E F E E G M V E G R A R G E R L A A L G
CAAGCAAACCAGCTTCTCTGGCAGCGTGGAGGTCATCGAAGTATCGTGACCCTTCAGAAGTCCCTGTGCCCTTGCTCCAGCCAGGCCAGGTATAAATATATATTATATATAAAACACACA 2063
K Q T S F S G S V E V I E V S *
GAAAAGGTAAATGGTTTTACTGCAATTTTTATCAAGAAGTAAATATTTCGATTTTTTATTTATTTAAGCTAGTGATCTGGCAACTGTGCGGGGCGGCCCTAAAGCTCTGTTTTTTACTGTC 2183
TGGTATTTAAACTGAAACAGGTTTCTAAGCAATATGAGGCCACCTTCAATCCCAAACTGGGTTGACAGGCCTGGGCCCCTCCTTGCCCCTCCCCTCTGGAAACATTACTGACCTTTCAAA 2303
GAGCTGCCCAGCTTTCCTGCACTTTTTACATAAGAAAAAAGGGGGGGGGGAA 2356

FIGURE 2

FIGURE 3

SmaI subclone of HB5a/3 #2 sequenced with T3 17/7/95

```
SCORES          Initl:   463 Initn:   527 Opt:   557
          85.2% identity in 182 bp overlap

    1049      1039      1029      1019      1009       999       990
M3-6.S AAGTGAGGTCCATCAGTCTGCAGGGCAGCCAGCAGCTTCAGACTCCTCTCATACTCCAGC
                                    |||||||||||| ||| ||| |||||||||
Hb5s2t                              AGCAGCTTCAGGCTCGGCTCGTACTCCAGC
                                          10        20        30

     989       979       969       959       949       939       930
M3-6.S AACTGGCCCAGGAAGTTGAAGTTGGGCGAGATGGAGGGGCGCCGATCCTTCAC^AACCTG
       | ||||||||||||||||||||||||||:|||||||  :::|:||  ||||||||  ||||||
Hb5s2t AGCTGGCCCAGGAAGTTGAAGTTGGGCNAGATGGACNNNCNCCTGTCCTTCACGAACCTG
               40        50        60        70        80        90

     929       919       909       899       889       879       870
M3-6.S TATGCGTCGTCAGAAGACATGCCCATGGTTTTCATGATGTACGCGATGGCAATGGTGGCA
       || ||:|||||| || |||||||||||||| |||||||||| |||||:|| ||||||||
Hb5s2t TAGGCNTCGTCGGAGGACATGCCCATGGTCTTCATGATGTAGGCGATNGCGATGGTGGC-
              100       110       120       130       140

     869       859       849       839       829       819       810
M3-6.S GAGCGAGAGATGCCAGCCAGACAGTGAACAATGACTTGGCAGCTGGACAGCTTGGCTTTA
       ||||: ||||||||||||||||||| ||:||
Hb5s2t GAGCNGGAGATGCCAGCCAGACAGTGGACNAT
     150       160       170       180

     809       799       789       779       769       759       750
M3-6.S TCAATAAACTCGATGGACTTGTCCAGCCAGGGCAGCAGCTTTTCACAGTAGTTGTCATTG


SCORES          Initl:   343 Initn:   343 Opt:   387
          78.8% identity in 156 bp overlap

            1880      1890      1900      1910      1920      1930
M3-6.S CGGAGGCGGGATGTGCGGACCGGCTGGCCCGAGGAGCCTGCTGCAGATGCACAGTTCAAG
                                    |||||||| || | || | ||||||||||
Hb5s1t                              GAGGAGCCGGCCCCGGAGACGCAGTTCAAG
                                          10        20        30

            1940      1950      1960      1970      1980      1990
M3-6.S AGGCGCAGCTGCCAGATGGAGTTCGAAGAGGGCATGGTGGAGGGGCGGCACGTGCAGAGC
       | |||||||||||||||||||||| |||||||||||||||||| ||:|: :  |||
Hb5s1t CGCCGCAGCTGCCAGATGGAGTTCGAGGAGGGCATGGTGGAGGGGC-GCNCNCNGCGAGG
              40        50        60        70        80

            2000      2010      2020      2030      2040      2050
M3-6.S TCCTGGCAGNCCT-GGCAAGCAAACCAGCTTCTCTGGCAGCGTGGAGGTCATCGAAGTAT
       |||||  |:||| |||||||  | |||||||| ||| ||||||||||||||| || |
Hb5s1t AGCTGGCCGCCCTGGGCAAGCTGGCGAGCTTCTCGGGCTGCGTGGAGGTCATCGAGGTGT
     90        100       110       120       130       140

            2060      2070      2080      2090      2100      2110
M3-6.S CGTGCCTTCAGAAGTCCCTGTGCCCTTGCTCCAGCCAGGCCAGGTATAAATATATATTAT
       | || |
Hb5s1t CCTGACCCC
     150
```

```
SCORES          Initl:   241 Initn:   241 Opt:   323
          82.4% identity in 119 bp overlap

     1359      1349      1339      1329      1319      1309
M3-6.S CNGGTNGGGCCGGGAAAGTCGGGCCTCCTGCTGGGTGCATAGGCCGACTTGATGTCCAGG
                                    || || ||   ||| ||||||||||||||||
Hb5s3t                              CTAGGGGCGATGGCAGACTTGATGTCCAGG
                                          10        20        30

     1299      1289      1279      1269      1259      1249
M3-6.S GAAAAGGAACGCTTGAGGCGGTTGGTGTCCTGGAGGCGGTCAGAGGAGAGGTGCAGGCCA
       || |||||| ||||||||||||||| |||||||| |||||||| |||||||||||| ||||
Hb5s3t GAGAAGGAGCGCTTGAGGCGGTTAGTGTCCTGCAGGCGGTCCGAGGAGAGGTGCTGGCCG
               40        50        60        70        80        90

     1239      1229      1219      1209      1199      1189
M3-6.S CGCAGGCCTGCTGCAGCCGCGCTGGTGGCTGGAGCTGTGCTGGGGATCGGAGCATCCCTC
       |||| |||:|||||| :|:|||||| :|
Hb5s3t CGCA-GCCTNCTGCAG-TNCNCTGGTCNC
               100       110


PstI subclone of HB5a/3 #2 sequenced with T7 primer 17/7/95

SCORES          Initl:   213 Initn:   213 Opt:   227
          79.5% identity in 78 bp overlap

       1029      1019      1009       999       989       979
M3-6.S CTGCAGGGCAGCCAGCAGCTTCAGACTCCTCTCATACTCCAGCAACTGGCCCAGGAAGTT
                                    ||| |||||||||| ||||||||||||||||
Hb5p2t                              CTCGTACTCCAGCAGCTGGCCCAGGAAGTT
                                          10        20        30

        969       959       949       939       929       919
M3-6.S GAAGTTGGGCGAGATGGAGGGGCGCCGATCCTTCACAAACCTGTATGCGTCGTCAGAAGA
       |||||||:::::|||||||| :::|||| |||||:|| |||||||| ||
Hb5p2t GAAGTTNNNNNAGATGGACNNNCGCCTGTCCTTNACGAACCTGTAGGC
               40        50        60        70

        909       899       889       879       869       859
M3-6.S CATGCCCATGGTTTTCATGATGTACGCGATGGCAATGGTGGCAGAGCGAGAGATGCCAGC


SCORES          Initl:   169 Initn:   169 Opt:   202
          81.3% identity in 75 bp overlap

     1330      1340      1350      1360      1370      1380
M3-6.S TTTCCCGGCCCNACCNGACCCCCGGTGAAGCCCCAAGCTCT-NAAGCT-GACAGCCNGTC
                                    ||| |||||||| :|||||| ||||||||:|||
Hb5s4t                              CCCCGAAGCTCTGCAAGCTGGACAGCCCGTC
                                          10        20        30
```

```
        1390       1400       1410       1420       1430       1440
M3-6.S TGGGGGNACACTGGGCCTGCCCTCGCCCAGCCCAGACAGCCCGGACTCCGTTCCAGAGTG
       ||||| : |:||||:|||| :||||||||||||| ||||||||||
Hb5s4t GGGGGCCCGCNCTGGNCCTGTNCTCGCCCAGCCCGGACAGCCCGGA
              40         50         60         70

        1450       1460       1470       1480       1490       1500
M3-6.S CCGCCCACGACCCCGCCGCGACGCCCCCCCGGCTAGTTCGCCTGCCCGCTCCCCCGCGCA
```

PstI subclone of HB5a/3 #2 sequenced with T3 primer 17/7/95

SCORES Initl: 151 Initn: 151 Opt: 241
76.0% identity in 104 bp overlap

```
        260        270        280        290        300        310
M3-6.S CTGGTGAAGCGGCGCCTTCAGCAGGGAAAAGTGACAATTGCTGAGCTTATCCAGCCTGCT
                                        ||| | ||||| |:::: ||||||||| |||
Hb5p2t                              TTCCGGGGTGGCCATTGCGGNNNNCATCCAGCCGGCT
                                        10         20         30

        320        330        340        350        360        370
M3-6.S ACACGGAGCCAGGTGGATGCCACAGAACCACAGGATGTAGTGGTGTATGACCAGAGCACA
       |||| || |||||||| || || || ||||||||| || ||||| ||||||||||||||
Hb5p2t GCACGCAG-CAGGTGGAGGCTACGGAGCCACAGGACGTGGTGGTCTATGACCAGAGCACG
       40         50         60         70         80         90

        380        390        400        410        420        430
M3-6.S CGAGATGCCAGCGTGCTGGCAGCAGACAGCTTCCTGTCCATCCTGCTCAGCAAGCTGGAC
       || ||  :||||:|
Hb5p2t CGGGA-CNCAGCNT
       100
```

SmaI subclone of HB5a/3 #2 sequenced with T7 17/7/95

SCORES Initl: 151 Initn: 151 Opt: 236
77.9% identity in 95 bp overlap

```
        260        270        280        290        300        310
M3-6.S CTGGTGAAGCGGCGCCTTCAGCAGGGAAAAGTGACAATTGCTGAGCTTATCCAGCCTGCT
                                        ||| | ||||| |:::: ||||||||| |||
Hb5s2t                                GGGTGGCCATTGCGGNNNNCATCCAGCCGGCT
                                           10         20         30

        320        330        340        350        360        370
M3-6.S ACACGGAGCCAGGTGGATGCCACAGAACCACAGGATGTAGTGGTGTATGACCAGAGCACA
       |||| || |||||||| || || || ||||||||| || ||||| ||||||||||||||
Hb5s2t GCACGCAG-CAGGTGGAGGCTACGGAGCCACAGGACGTGGTGGTCTATGACCAGAGCACG
              40         50         60         70         80         90

        380        390        400        410        420        430
M3-6.S CGAGATGCCAGCGTGCTGGCAGCAGACAGCTTCCTGTCCATCCTGCTCAGCAAGCTGGAC
       || ||
Hb5s2t CGGGA
```

```
SCORES          Initl:   133 Initn:   133 Opt:   224
          69.8% identity in 116 bp overlap

     1240      1250      1260      1270      1280      1290
M3-6.S CCTCTGACCGCCTCCAGGACACCAACCGCCTCAAGCGTTCCTTTTCCCTGGACATCAAGT
                                   ||||||| ||||| ||||||||||||||||
Hb5p1t                             TCAAGCG-TCCTTCTCCCTGGACATCAAGT
                                          10        20

     1300      1310      1320      1330      1340      1350
M3-6.S CGGCCTATGCACCCAGCAGGAGGCCCGACTTTCCCGGCCCNACCNGACCCCCGGTGAAGC
       | |||||| :| || ||||||| :|||||||::::||::::|   ||:|||||||:: :    |
Hb5p1t CTGCCTACNCCCCTAGCAGGGNGCCCGANNNNCCNNNNC---CCCGACCCCNNCNAGGCC
      30        40        50        60            70        80

     1360      1370      1380      1390      1400      1410
M3-6.S CCCAAGCTCT--NAAGCT-GACAGCCNGTCTGGGGGNACACTGGGCCTGCCCTCGCCCAG
       || ||||||| ||||||:|||::|:
Hb5p1t CCGAAGCTCTCGNAAGCTGGACANNCGTCG
        90        100       110
```

```
SCORES          Initl:   129 Initn:   129 Opt:   213
          72.0% identity in 100 bp overlap

          1000      1010      1020      1030      1040      1050
M3-6.S CTGAAGCTGCTGGCTGCCCTGCAGACTGATGGACCTCACTTGGGGACCCCTGAGCCCCTC
                                  || | | ||  |||||:|: ||||| |
Hb5p3t                          GACCCGGGCACCCCCTCAGGGACNCNGGAGCCTCCG
                                      10        20        30

          1060      1070      1080      1090      1100      1110
M3-6.S ATGGGCCCGGCAGCAGGCATCCCACTGCCCCGGCTGCCACCATCTACCTCAGAGAGCGCT
       | || || || ||   ||| ||||| |||||||||||| |||||||||||||:||
Hb5p3t CCCAGTCCTGCCGCCGGGGCCCCGCTGCCACGGCTGCCACCACCTACCTCAGAGAGCNCT
        40        50        60        70        80        90

          1120      1130      1140      1150      1160      1170
M3-6.S GCCACTGGGAGCGAGGCAGCCACCGCAGCCAGGGAGGGCAGCCCAAGTGCTGGAGGGATG
       ||||| ||||
Hb5p3t GCCACAGGGAATNCNNCTNCAGGGAGG
        100       110       120
```

```
SCORES          Initl:   129 Initn:   129 Opt:   196
          66.9% identity in 127 bp overlap

     1239      1229      1219      1209      1199      1189

M3-6.S CGCAGGCCTGCTGCAGCCGCGCTGGTGGCTGGAGCTGTGCTGGGGATCGGAGCATCCCTC
                                        || |||| |  ||| :  || |  ||||
Hb5p3t                                GGGGGCTGGG--GGGCNCGGGGGGCTCCCCG
                                            10          20

     1179      1169      1159      1149      1139      1129
M3-6.S CAGCACTTGGGCTGCCCTCCCTGGCTGCGGTGGCTGCCTCGCTCCCAGTGGCAGCGCTCT
       | :| :|  ||| |||||||||          |||| ||| |  |||| |||||:|||:||||
Hb5p3t CCNC-NTCAGGCCGCCCTCCC---------TGGCAGCCGCATTCCCTGTGGCAGCNCTCT
      30         40                  50        60        70

     1119      1109      1099      1089      1079      1069
M3-6.S CTGAGGTAGATGGTGGCAGCCGGGGCAGTGGGATGCCTGCTGCCGGGCCCATGAGGGGCT
       |||||||| |||||||||||| ||||| :||:  ||
Hb5p3t CTGAGGTAGGTGGTGGCAGCCGTGGCAGCNGGNCCCC
      80        90       100       110

     1059      1049      1039      1029      1019      1009
M3-6.S CAGGGGTCCCCAAGTGAGGTCCATCAGTCTGCAGGGCAGCCAGCAGCTTCAGACTCCTCT
```

```
SCORES          Initl:   112 Initn:   112 Opt:   184
          83.1% identity in 65 bp overlap

           1609      1599      1589      1579      1569      1559
M3-6.S GCCGGCTGGCAGGGCCAGGCAGCCCGGGCGCCGACAGGGCCGAGA-GSSTGCCGTGGAGT
                                    |||||||||||||||  ::|||||  |||||
Hb5s4t                              CCGACAGGGCCGAGACCGGTGCCGCGGAGT
                                        10        20        30

           1549      1539      1529      1519      1509      1499
M3-6.S CTGCCGGGCCGTGTCTCCAAAGTTCAGGCCCAGACCATGCGCGGGGGAGCGGGCAGGCGA
       ||||||||||| || || ||||||||||| ||| |
Hb5s4t CTGCCGGGCCGCATCGCCGAAGTTCAGGCCGAGGCTGT
           40        50        60

           1489      1479      1469      1459      1449      1439
M3-6.S ACTAGCCGGGGGGGCGTCGCGGCGGGGTCGTGGGCGGCACTCTGGAACGGAGTCCGGGCT
```

```
SCORES          Initl:    67 Initn:    67 Opt:    91
          58.6% identity in 87 bp overlap

            1310      1320      1330      1340      1350
M3-6.S GCCTATGCACCCAGCAGGAGGCCCGACTTTCCCGGCCCNACCNGACCCCCGG---TGAAG
                                  |||| |||:   :  ||| |||   |||||
Hb5s5t              CCAAACAGCACCCCGCCCGCCCCCTGTGCGCCCTCGGGGCTGAAG
                               10        20        30        40

     1360      1370      1380      1390      1400      1410
M3-6.S CCCCAAGCTCTNAAGCTGACAGCCNGTCTGGGG-GNACACTGGGCCTGCCCTCGCCCAGC
       | ||| |  | :  || ||| || :| |||||| |:  |  ||   |||||  ||||
Hb5s5t CACCAGGGCCCGTCGCGGACGGCGTGCCTGGGGAGTCGAGCGGCGGTGCCCAGGCCCCAT
         50        60        70        80        90       100

     1420      1430      1440      1450      1460      1470
M3-6.S CCAGACAGCCCGGACTCCGTTCCAGAGTGCCGCCCACGACCCCGCCGCGACGCCCCCCCG

Hb5s5t C
```

SmaI subclone of HB5a/3 #2 sequenced with T3 17/7/95

```
SCORES          Initl:    44 Initn:    44 Opt:    62
          53.5% identity in 114 bp overlap

    250       260       270       280       290       300
M3-6.S TTCAAAGCTGGTGAAGCGGCGCCTTCAGCAGGGAAAAGTGACAAT-TGCTGAGCTTATCC
                                  || :| |  |||::: : |||  |||  |
Hb5s2t TCCAGCAGCTGGCCCAGGAAGTTGAAGTTGGGCNAGATGGACNNNCNCCTGTCCTTCACG
         30        40        50        60        70        80

     310       320       330       340       350       360
M3-6.S AGCCTGCTACACGGAGCCAGGTGG--ATGCCACAGAACCACAGGATGTAGTGGTGTATGA
       | ||||| ||  |:  | | || ||  |||||  |  |  || |||||||  |  :  ||
Hb5s2t AACCTG-TAGGCNTCGTC-GGAGGACATGCCCATGGTCTTCATGATGTAGGCGATNGCGA
         90        100       110       120       130       140

         370       380       390       400       410       420
M3-6.S --CCAGAGCACACGAGATGCCAGCGTGCTGGCAGCAGACAGCTTCCTGTCCATCCTGCTC
           | |  |: ||||||||||
Hb5s2t TGGTGGCGAGCNGGAGATGCCAGCCAGACAGTGGACNAT
           150       160       170       180
```

FIGURE 4

FIGURE 5

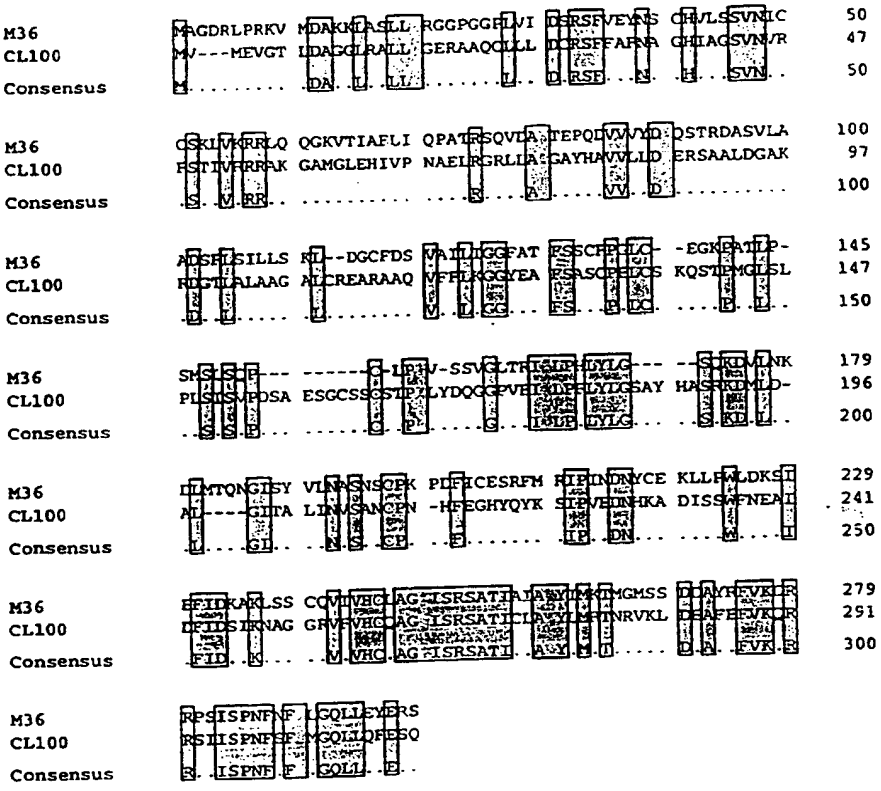
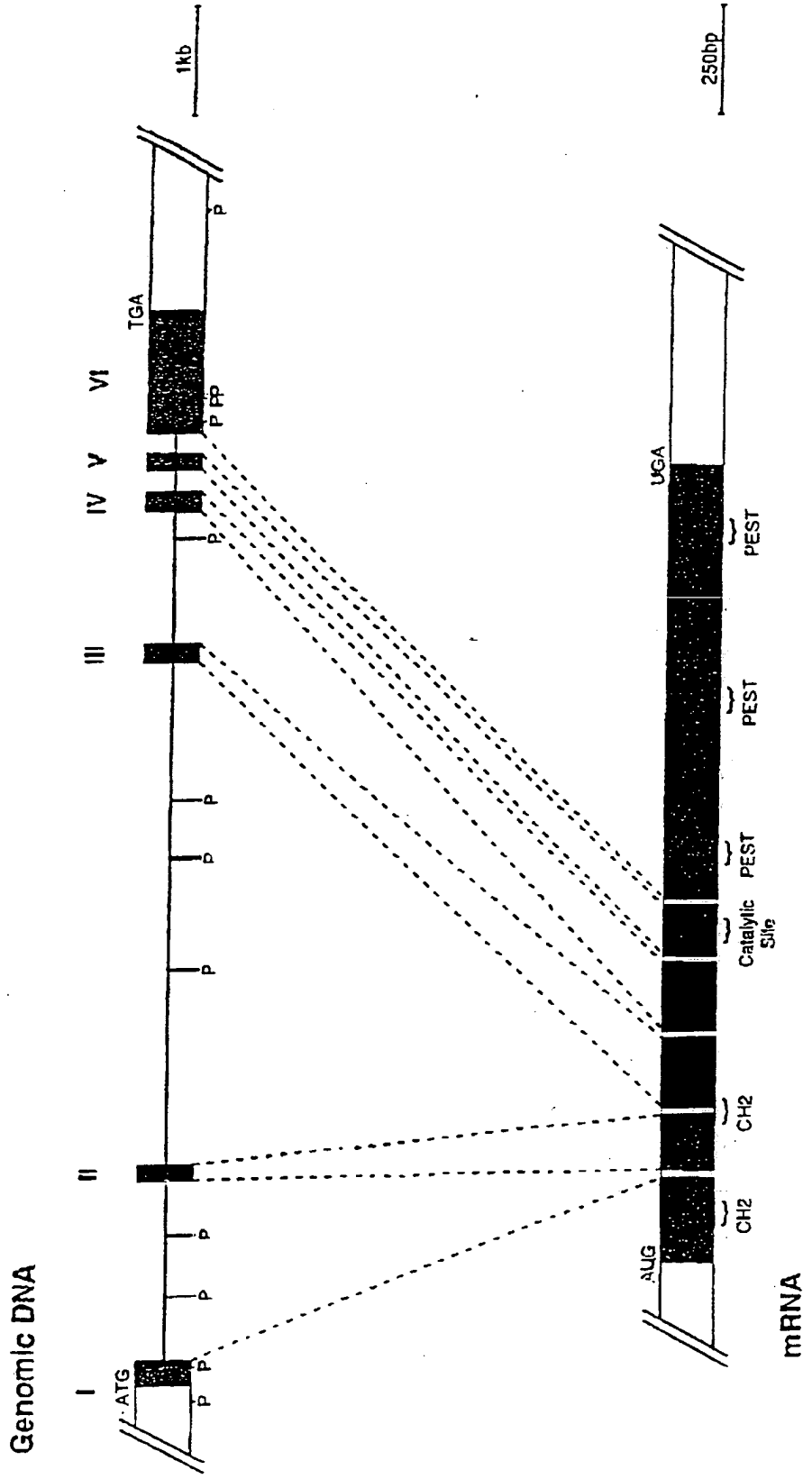


FIGURE 6


1kb
250bp
TGA
VI
V
IV
III
II
I
ATG
P
UGA
PEST
PEST
PEST
Catalytic Site
CH2
CH2
AUG
Genomic DNA
mRNA

# INTERNATIONAL SEARCH REPORT

Internatior. pplication No
PCT/GB 96/01906

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | JOURNAL OF NEUROCHEMISTRY, vol. 65, no. 4, 1995, PHILADELPHIA,US, pages 1823-1833, XP000196676 K.J. MARTELL ET AL.: "hVH-5: a protein-tyrosine phosphatase abundant in brain that inactivates mitogen-activated protein kinase" cited in the application see the whole document | 1-4 |
| A | --- -/-- | 5-23 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 14 January 1997 | 29. 01. 97 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. ( + 31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax ( + 31-70) 340-3016 | Mateo Rosell, A.M. |

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Internation. pplication No
PCT/GB 96/01906

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | JOURNAL OF CELL SCIENCE, vol. 108, no. 8, 1995, CAMBRIDGE, GB, pages 2885-2896, XP000196666 T.A. LEWIS ET AL.,: "XCL100, an inducible nuclear MAP kinase phosphatase from Xenopus laevis: its role in MAP kinase inactivation in differentiated cells and its expression during early development" see the whole document | 1 |
| A | | 2,5,6, 10,11, 15-18, 21-23 |
| | --- | |
| X | ONCOGENE, vol. 7, no. 1, 1992, HAMPSHIRE,GB , pages 187-190, XP000196653 C.H. CHARLES: "cDNA sequence of a growth-factor inducible immediate early gene and characterization of its encoded protein" cited in the application see the whole document | 1 |
| A | | 2,5,6, 10,11, 15-18, 21-23 |
| | --- | |
| X | JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 269, no. 5, 1994, MARYLAND, US, pages 3596-3604, XP000611984 S.P. KWAK ET AL.: "Isolation and characterization of a human dual specificity protein-tyrosine phosphatase gene" cited in the application | 1 |
| A | see the whole document | 2,5,6, 10,11, 15-18, 21-23 |
| | --- | |
| A | BIOCHIMICA ET BIOPHISICA ACTA, vol. 1256, no. 2-3, 1995, AMSTERDAM, NL, pages 152-160, XP000196716 S.M. KEYSE: "An emerging family of dual specificity MAP kinase phosphatases" see the whole document | 1,5, 21-23 |
| | --- | |
| A | WO,A,94 01119 (NATIONAL UNIVERSITY OF SINGAPORE ) 20 January 1994 see page 2, line 8-15 see page 8, line 13-15 | 16 |
| | --- | |
| | -/-- | |

2

# INTERNATIONAL SEARCH REPORT

International application No
PCT/GB 96/01906

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | WO,A,95 06735 (THE LUDWIG INSTITUTE FOR CANCER RESEARCH ) 9 March 1995<br>see page 43 - page 44; examples 5,AND,8 | 16-18 |
| P,X | JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 8, 1996, MARYLAND, US, pages 4319-4326, XP000611987<br>M. MUDA ET AL.: "MKP-3 a novel cytosolic protein-tyrosine phosphatase that exemplifies a new class of mitogen activated kinase phosphatase" | 1 |
| A | see the whole document | 2,5,6, 10,11, 15-18, 21-23 |
| P,X | HUMAN MOLECULAR GENETICS, vol. 5, no. 5, 1996, OXFORD, GB, pages 675-684, XP000611971<br>A.M. THEODOSIOU: "A member of the MAP kinase phosphatase gene family in mouse containing a complex trinucleotide repeat in the coding region"<br>cited in the application<br>see the whole document | 1-15 |
| P,A | WO,A,95 30774 (BECKMAN INSTRUMENTS INC ) 16 November 1995<br>see page 1, line 1-6<br>see page 7, line 9 - page 8, line 35 | 19,20 |

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Internation. ..plication No
PCT/GB 96/01906

| Patent document cited in search report | Publication date | Patent family member(s) | | Publication date |
|---|---|---|---|---|
| WO-A-9401119 | 20-01-94 | AT-T- | 143808 | 15-10-96 |
| | | DE-D- | 69305348 | 14-11-96 |
| | | EP-A- | 0648123 | 19-04-95 |
| WO-A-9506735 | 09-03-95 | AU-A- | 7644394 | 22-03-95 |
| | | CA-A- | 2170515 | 09-03-95 |
| WO-A-9530774 | 16-11-95 | AU-A- | 2360195 | 29-11-95 |